European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 07, Issue 02, 2020

Multispectral Image Classification Using Neural Networks in Astronomical Imagery

A. Praveena[1], Nimisha Asthana[2], Anusha Chattopadhyay[3], Yash Singh[4]

[1]*A.P CSE, Department of Computer Science Engineering, SRM Institute of Technology, Chennai, Tamil Nadu, India.*

[2]*Department of Computer Science Engineering, SRM Institute of Technology, Chennai, Tamil Nadu, India.*

[3]*Department of Computer Science Engineering, SRM Institute of Technology, Chennai, Tamil Nadu, India.*

[4]*Department of Computer Science Engineering, SRM Institute of Technology, Chennai, Tamil Nadu, India.*

[1]*PraveenaSrmcse@gmail.com*

[2]*nimisha.asthana37@gmail.com*

[3]*anushachattopadhyay_s@srmuniv.edu.in*

[4]*yashsingh_v@srmuniv.edu.in*

Abstract: A multispectral image is an image that has wavelengths across the spectrum of electromagnetism. Astronomical images have various layers in their image capture process and thus fit into this category. This research aims to analyze the different lights of astronomical objects and their images and their colors. It focuses on how neural network models learn from each attribute of the images, thus aiming to find correlation and importance of the attributes of the images. Attempting to allow the CNN model to classify the images more efficiently by sorting the data.

Index Terms - Astronomy, Multispectral, CNN, SDSS, Galaxy Zoo

1. INTRODUCTION

The proposed system emphasizes the image displaying and showing a correlation between the different aspects of astronomical objects. Through breaking down the images we aim to find a correlation between the different aspects of the images. Data from the Sloan Digital Sky Survey [1] and Galaxy Zoo[2][3] has been used for the system and analysis. The procedure was to make models that work with our data and then break the data down and test the same models against it. This would hence give us a better approach to what data is more important to the machine. The astroNN[4] python package has helped us with this part. We also plotted the data independently to get a vision of what was going on with the magnitude of the data.

2. CLASSIFIERS

To analyze the data we used a basic classifier and decided to see how it fairs under our data manipulation. To do this we built a classifier to the best of our abilities for the existing data we had, and then worked with breaking down and taking away the data and seeing how the accuracy changes.

The Technology Used and Procedure

The models were built using Tensorflow with Keras and Sklearn. We also used astroNN for one of the analyses. We took different approaches based on the dataset type, as to build the best of what we can before we challenge our classifiers by breaking its data.

3. SDSS DATASET ANALYSIS

Introduction to the Dataset: Tabular Version

For this project, we have used the SDSS: Sloan Digital Sky Survey DR14 dataset. The SDSS DR14 dataset is the 2nd data release of the 4th phase of the Sloan Digital Sky Survey [3]. DR14 contains SDSS observations during July 2016.

Table 1: Snippet of the SDSS DR14 Dataset[5]

u	g	r	i	z	redshift	fiberid	class
19.4741	17.0424	15.947	15.5034	15.2253	-8.96E-06	491	STAR
18.6628	17.2145	16.6764	16.4892	16.3915	-0.0000549	541	STAR
19.383	18.1917	17.4743	17.0873	16.8013	0.123111	513	GALAXY
18.7383	18.6096	18.397	18.3117	17.9766	0.271937	587	QSO

Since we decided to focus on light and images, we looked into the U, G, R, I, Z and Redshift columns of the dataset.

U, G, R, I, Z (Ultraviolet, Visible and Infrared Spectrums) and Redshift Analysis

To analyze the focus of the light for the objects, STAR, GALAXY, and QSO, we plotted out the data of the dataset to look for any potential patterns in the wavelengths.

Here we plotted out histograms of the data to see the comparisons of U, G, R, I, Z spectrum measurements and the Redshift analysis consisting of STARS, GALAXY and QSO data.

Table 2: Midpoint of the Photometric Bands

Band	Effective Wavelength Midpoint(nm)
U	365
G	~475
R	358
I	806
Z	900

i) U-Band analysis- ULTRAVIOLET light

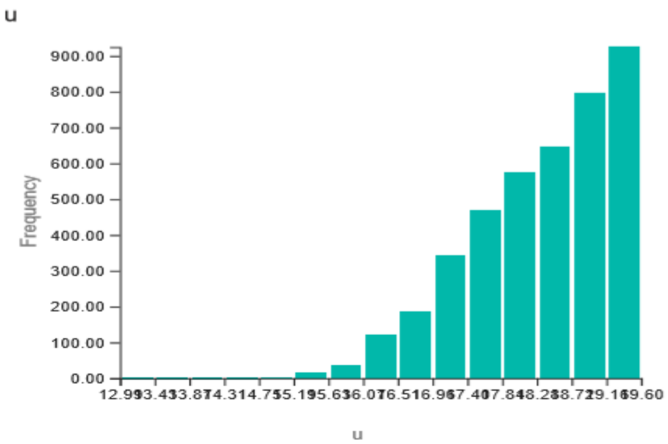

Fig. 1 U-band Plot of STARS

ISSN 2515-8260 Volume 07, Issue 02, 2020

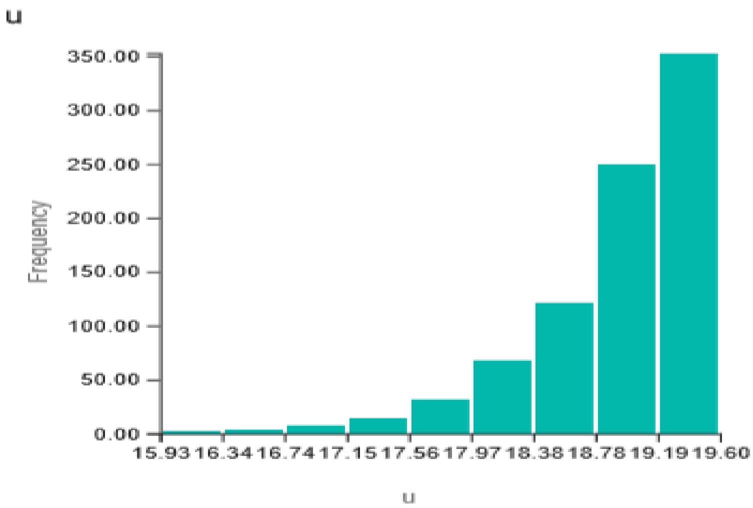

Fig. 2 U-band Plot of QUASARS

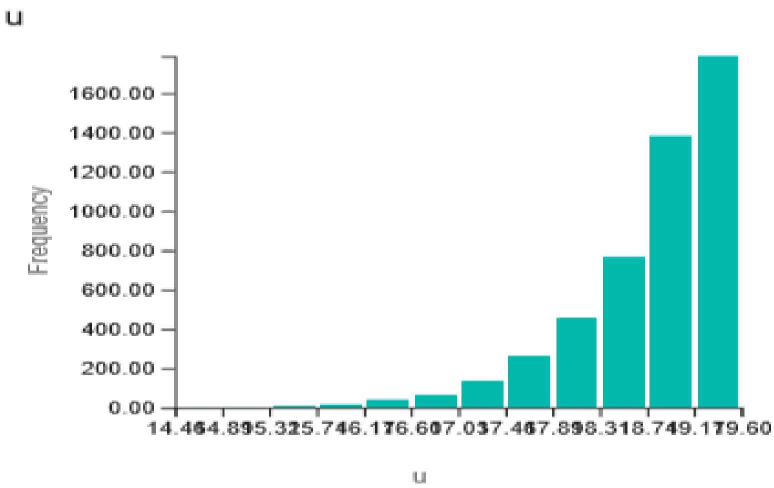

Fig. 3 U-band Plot of GALAXIES

ii) G-Band and R-band analysis- Visible light

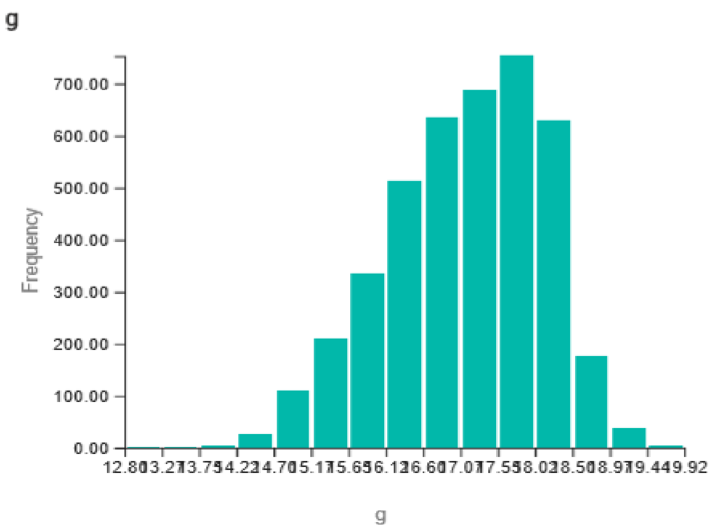

Fig. 4 G-band Plot of STARS

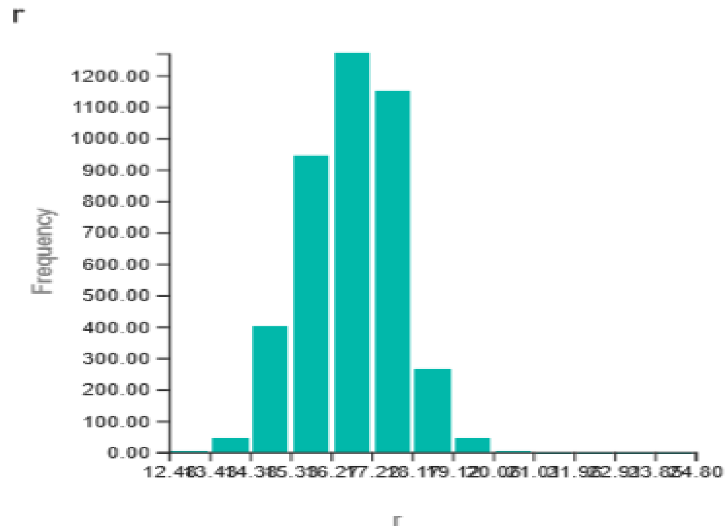

Fig. 5 R-band Plot of STARS

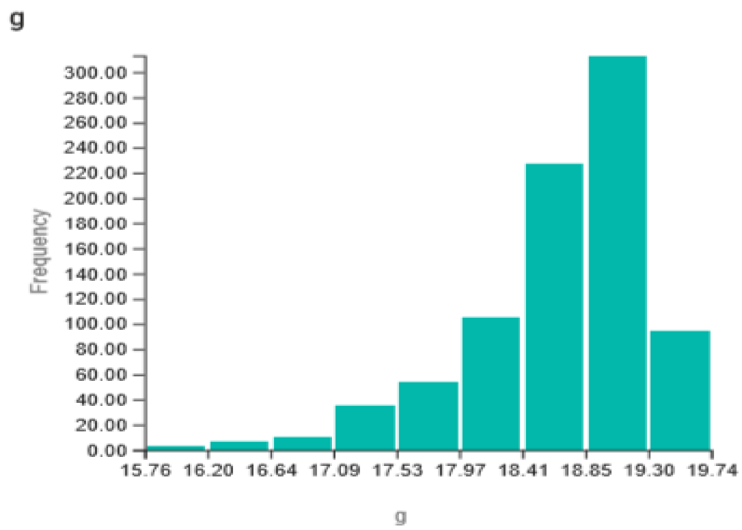

Fig. 6 G-band Plot of QUASARS

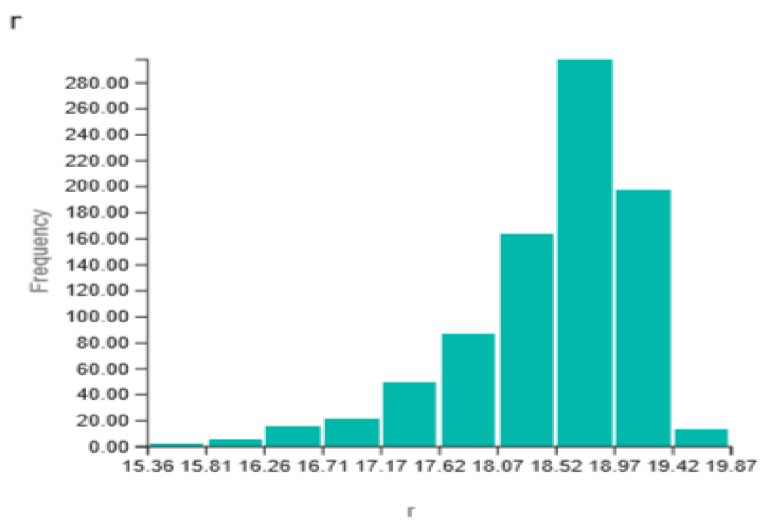

Fig. 7 R-band Plot of QUASARS

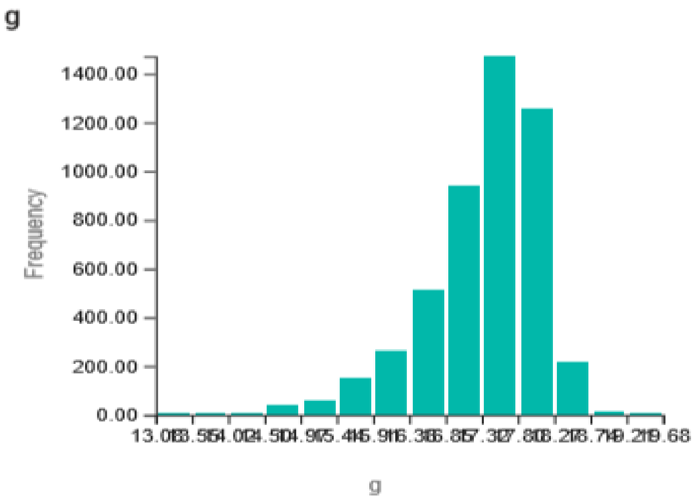

Fig. 8 G-band Plot of GALAXIES

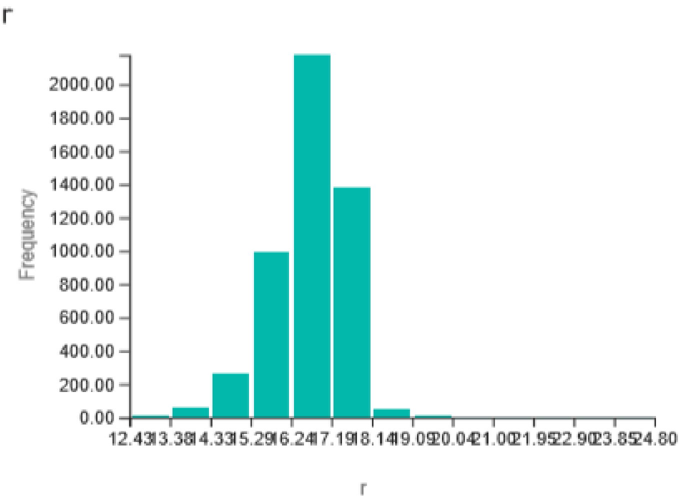

Fig. 9 R-band Plot of GALAXIES

iii) I-Band and Z-Band analysis- Infrared light

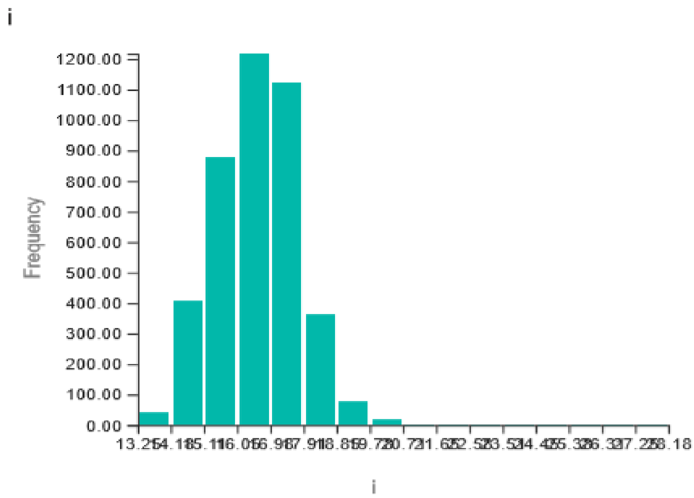

Fig. 10 I-band Plot of STARS

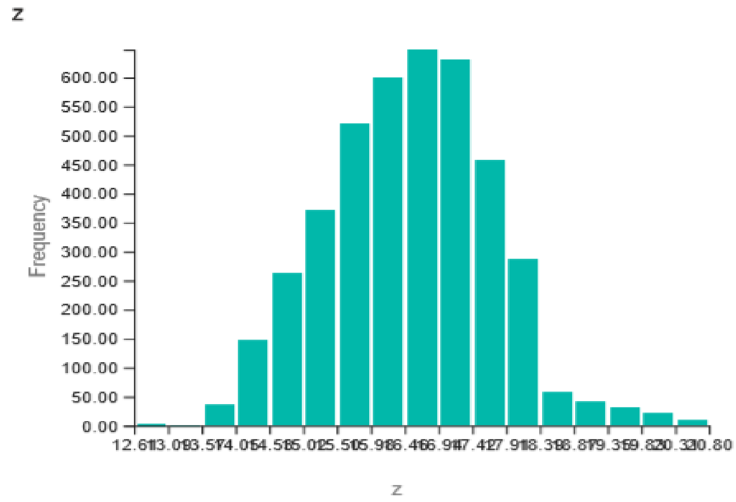

Fig. 11 Z-band Plot of STARS

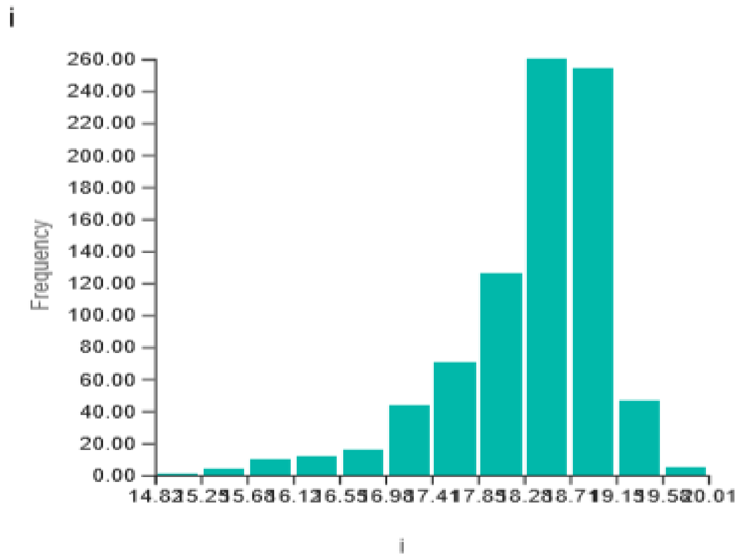

Fig. 12 I-band Plot of QUASARS

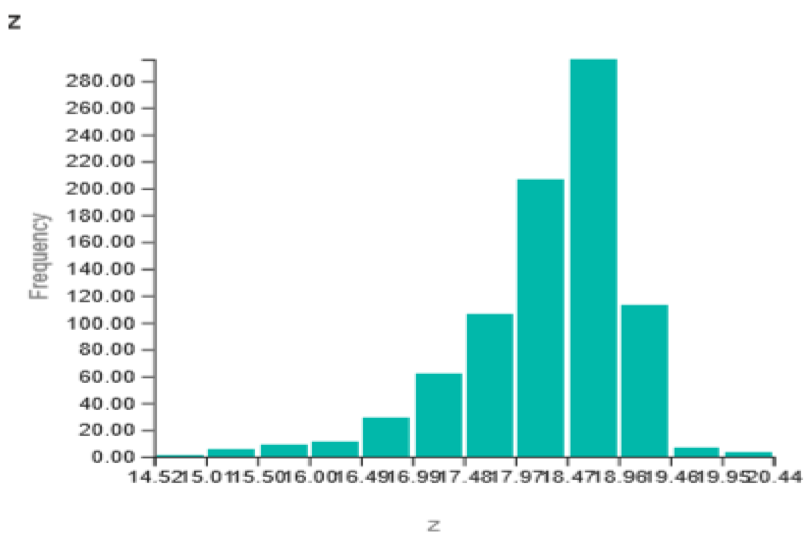

Fig. 13 Z-band Plot of QUASARS

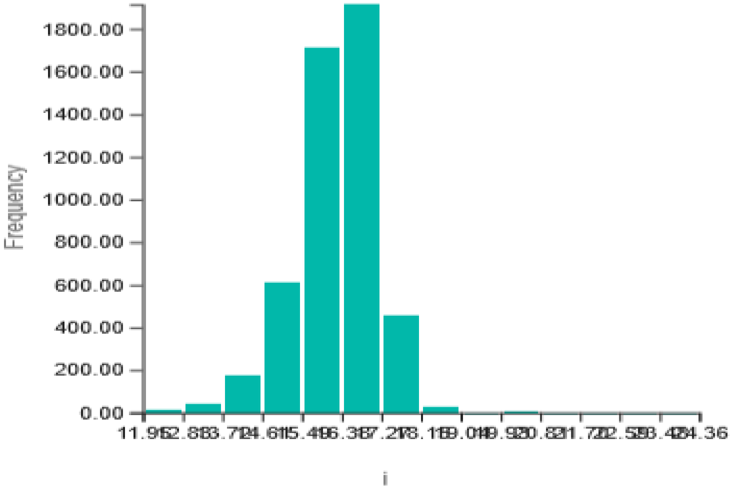

Fig. 14 I-band Plot of GALAXIES

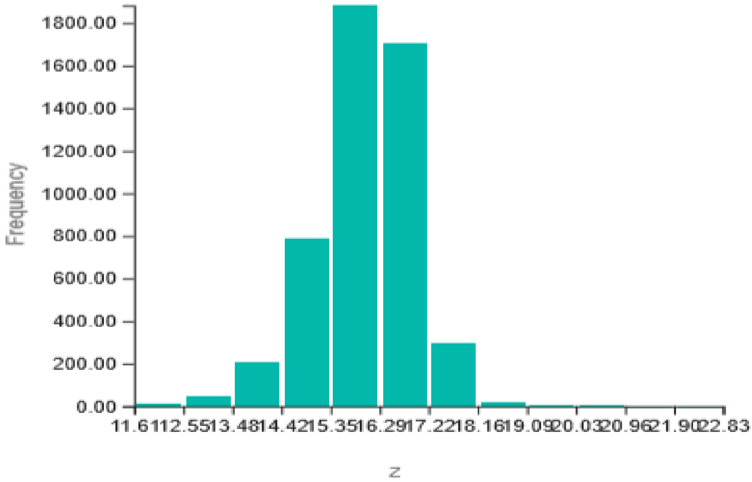

Fig. 15 Z-band Plot of GALAXIES

iv) Redshift analysis

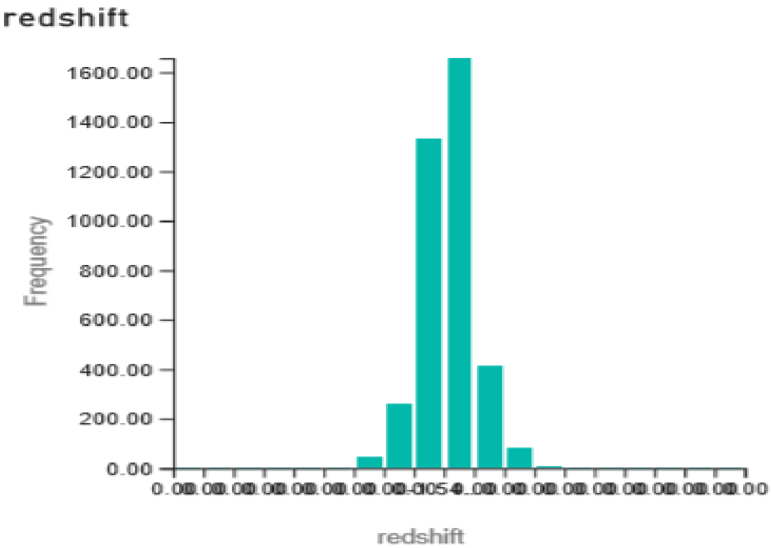

Fig. 16 Redshift plot of STARS

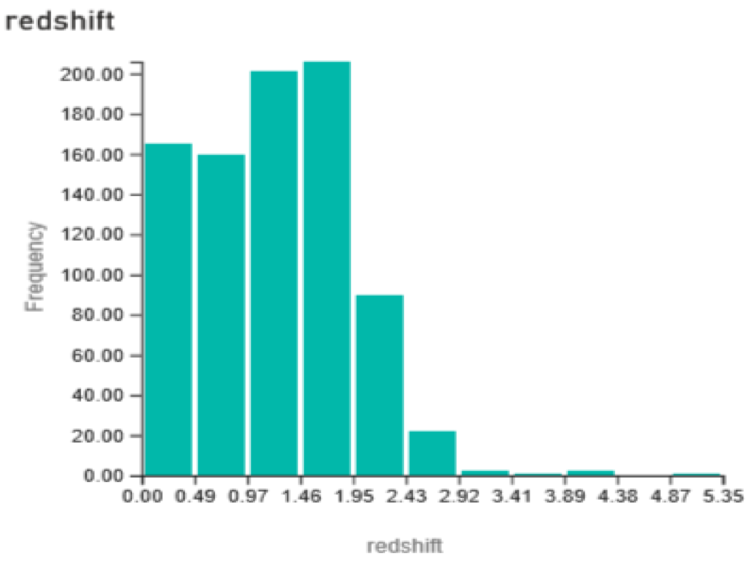

Fig. 17 Redshift plot of QUASARS

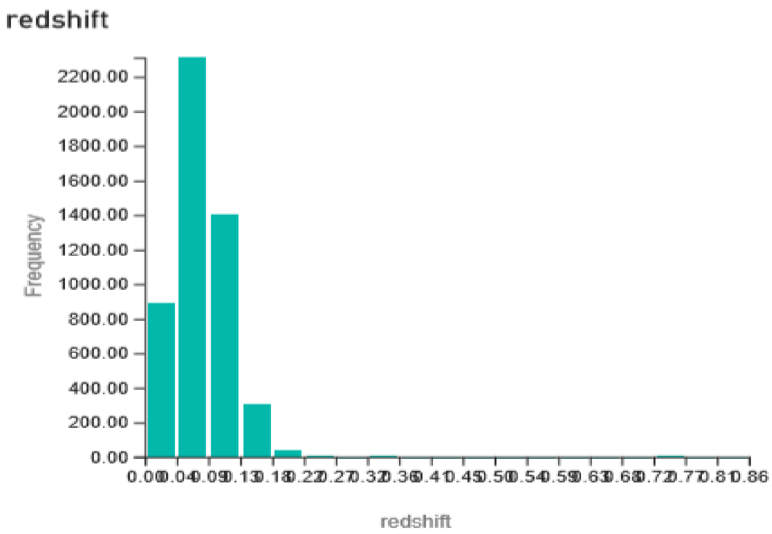

Fig. 18 Redshift Plot of GALAXIES

Upon analysis, we see that stars give more varied frequencies, whereas quasars and galaxies emit closer to a particular spectrum of ultraviolet and visible, infrared wavelengths. Quasars redshift is more varied as compared to stars and galaxies.

Prediction Analysis of the Data

We built a prediction model using sklearn and then use different comparisons and see how the accuracies compare

Test with Ultraviolet light: U band

	Precision	recall	f1-score	support
GALAXY	0.99	0.98	0.99	1514
QSO	0.94	0.92	0.93	236
STAR	0.99	1.00	1.00	1250
Accuracy			0.99	3000
Macro Average	0.98	0.97	0.97	3000
Weighted Average	0.99	0.99	0.99	3000

With the system only using U band for comparison, it gets a pretty good comparison what the object maybe

Test with Visible light: G, R bands

	precision	recall	f1-score	support
GALAXY	0.99	0.98	0.99	1514
QSO	0.94	0.91	0.93	236
STAR	0.99	1.00	1.00	1250
Accuracy			0.99	3000
Macro Average	0.97	0.97	0.97	3000
Weighted Average	0.99	0.99	0.99	3000

The precision and accuracy values do not change much.

Test with infrared light: I, Z bands

	Precision	recall	f1-score	support
GALAXY	0.99	0.99	0.99	1514
QSO	0.95	0.93	0.94	236
STAR	0.99	1.00	1.00	1250
Accuracy			0.99	3000
Macro Average	0.98	0.97	0.97	3000
Weighted Average	0.99	0.99	0.99	3000

The QSO precision seems to jump here, however only slightly.

Test with Redshift

	Precision	recall	f1-score	support
GALAXY	0.95	0.93	0.94	1514
QSO	0.89	0.89	0.89	236
STAR	0.93	0.95	0.94	1250
Accuracy			0.94	3000
Macro Average	0.92	0.92	0.92	3000
Weighted Average	0.94	0.94	0.94	3000

The accuracy jumps down for just redshift data. This might be because the frequencies vary a lot for redshift data, especially when we look at Quasar data.

4. SDSS AND GALAXY ZOO DATASET IMAGES ANALYSIS

Introduction to the Images

The images are from Galaxy zoo[2][3], which is a project that has put SDSS in the public domain to differentiate the shapes of galaxies with the help of the public. We have attempted to do so with astroNN; a python package meant for running various neural networks. As astroNN [4] kept the galaxy10 [4] database, we could use it directly for the SDSS[1][5] image analysis. Then we changed the attributes of the images to see the effects.

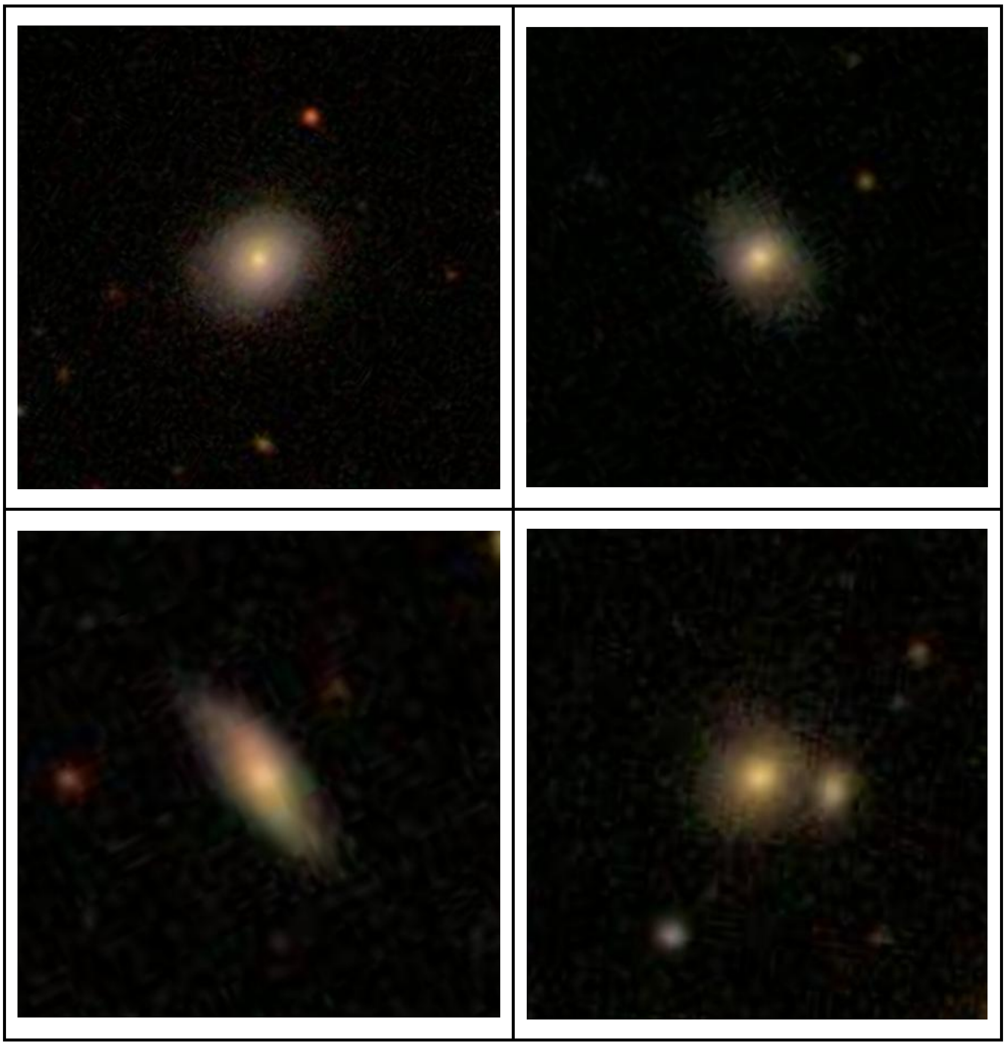

Fig. 19 Sample of the Images Part of the SDSS- Galaxy Zoo Data

The images are categorized into the following:
Class 0: Disk, Face-on, No Spiral
Class 1: Smooth, Completely round
Class 2: Smooth, in-between round
Class 3: Smooth, Cigar-shaped
Class 4: Disk, Edge-on, Rounded Bulge
Class 5: Disk, Edge-on, Boxy Bulge
Class 6: Disk, Edge-on, No Bulge
Class 7: Disk, Face-on, Tight Spiral
Class 8: Disk, Face-on, Medium Spiral
Class 9: Disk, Face-on, Loose Spiral
This is how the galaxy zoo has classified these images.

Analysis of the Images

We decided to take the images with high confidence and low confidence in morphology as described in the galaxy zoo challenge and then alter the images and see how it changes our results. astroNN already has the galaxy zoo as kept in the galaxy10 dataset. This thus helps us with our challenge. We decided to run original colored images at first and then run greyscale images.

Original results:

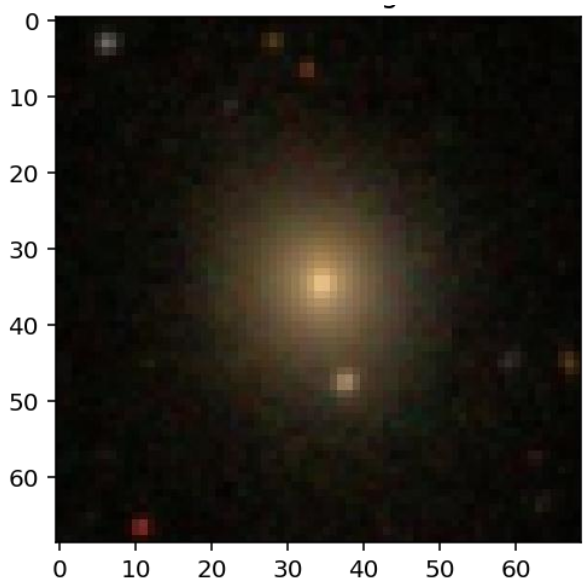

Fig. 20 Sample of the Unaltered Image

With the initial run of astroNN galaxy10 database, we have the following results:

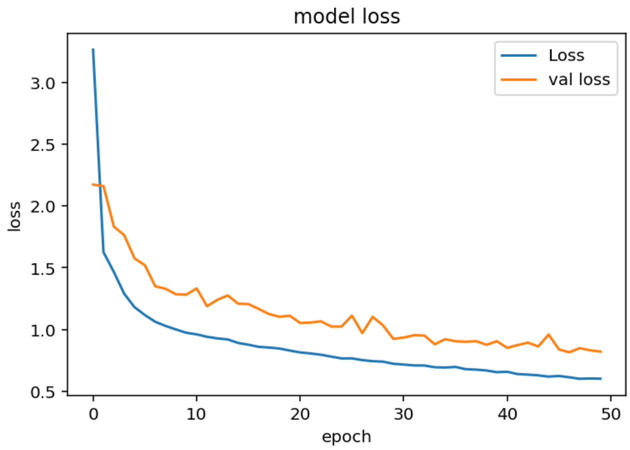

Fig. 21 Model Run with Unaltered Images: Loss

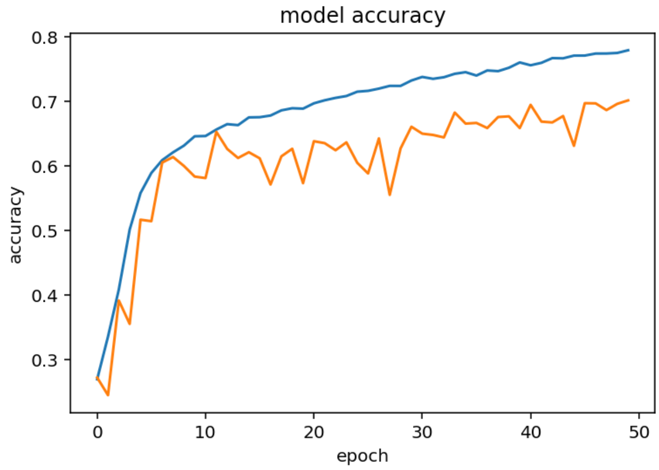

Fig. 22 Model Run with Unaltered Images: Accuracy

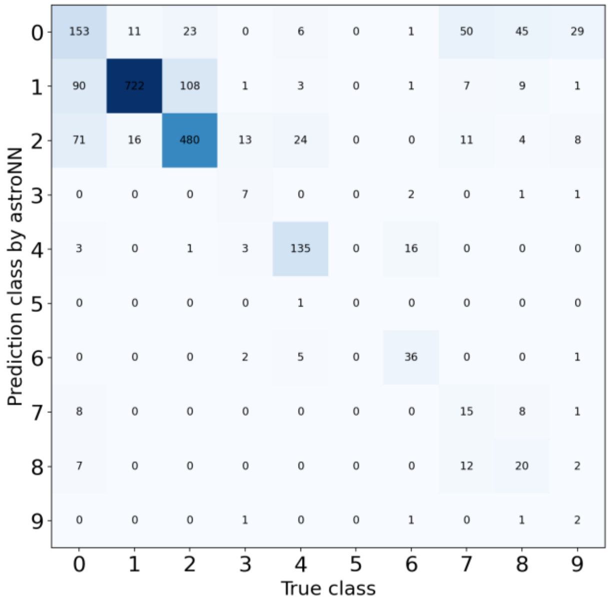

Fig. 23 Model Run with Unaltered Images: Confusion Matrix

Results when all images are filtered to grayscale:

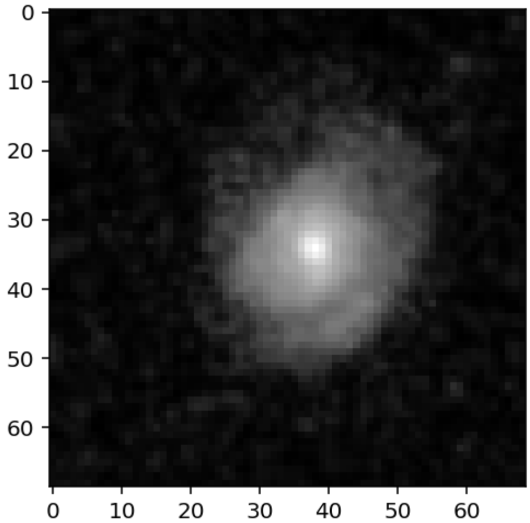

Fig. 24 Sample of Image Altered to Grayscale

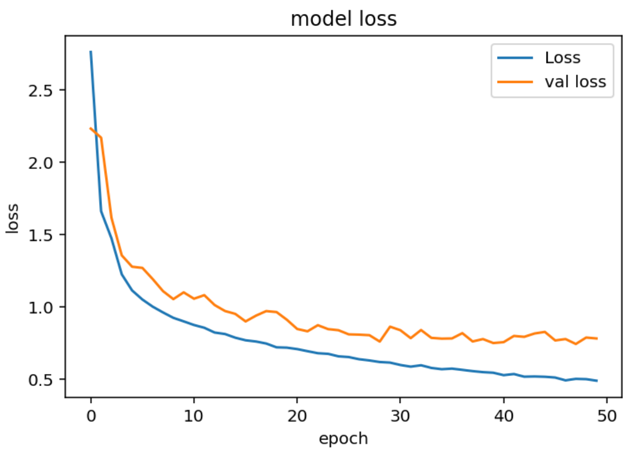

Fig. 25 Model Run with Altered Images to Grayscale: Loss

European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 07, Issue 02, 2020

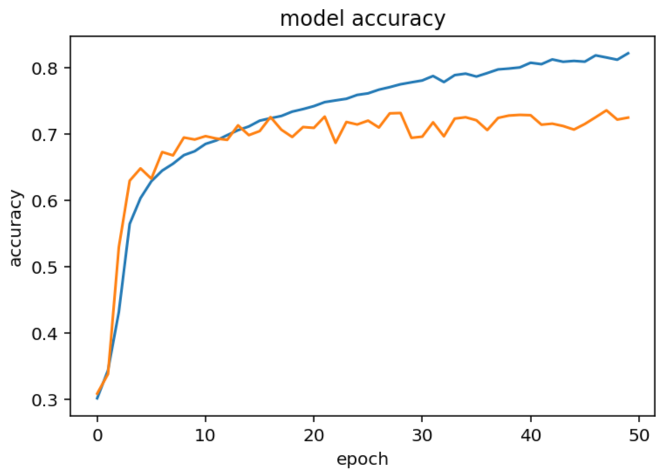

Fig. 26 Model Run with Altered Images to Grayscale: Accuracy

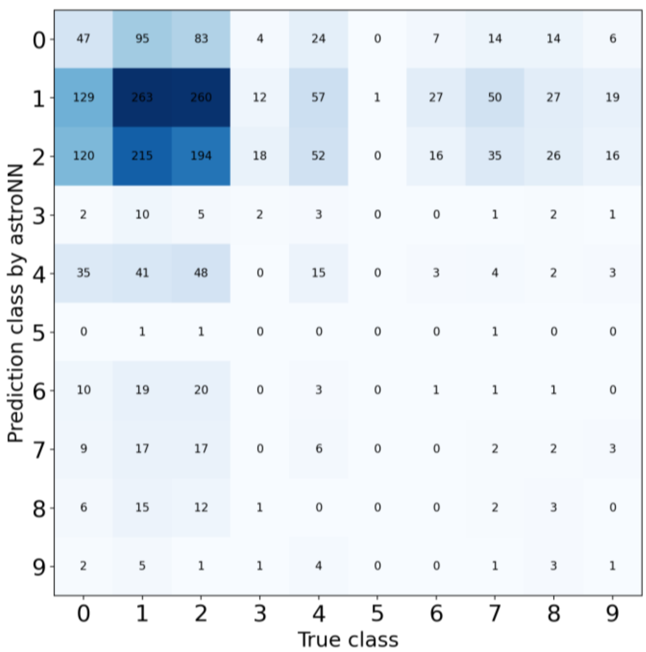

Fig. 27 Model Run with Altered Images to Grayscale: Confusion Matrix

Conclusion based on these results of image analysis

When we compare the two results of unaltered (fig 23) vs. greyscale (fig 27) we can see an increase of false positives when images turn to greyscale. Thus suggesting that color is an important factor while distinguishing the classes effectively.

5. CONCLUSION

After looking at the analysis of the two parts of the study, tabular analysis, and image analysis. We can see that color is an important factor when it comes to distinguishing astronomical images. It was stated in previous multispectral analysis' for normal images[6][7] that RGB spectrums play a major role in image analysis. We now see that when it comes to identifying the class of object (in case of tabular analysis) or when it comes to analyzing the shape of multiple objects of the same class (image analysis) we see that even false-colored images (such as images in space) colors are important factors for objective machine classification.

6. ACKNOWLEDGMENT

First, we would like to thank Mrs. Praveena for enlightening us with her knowledge and encouraging us with her positive feedback.

Also, we would like to thank Mr. Anirban Mukherjee for helping us with his expertise on neural networks, this project would not have been possible without you. Lastly, we are grateful to our parents for supporting us both mentally and financially.

7. REFERENCES

[1] Blanton et al. 2017 (SDSS-IV Overview), Bundy et al. 2015 (MaNGA technical description), Dawson et al. 2016 (eBOSS technical description), Majewski et al. 2017, AJ, 154, 94 (APOGEE Overview Paper), Yan et al. 2018, ApJ, accepted (MaStar Overview Paper), Eisenstein et al. 2011 (SDSS-III technical description), Dawson et al. 2013 (BOSS technical description), Yanny et al 2009, AJ, 137, 4377 (SEGUE Overview), Majewski et al 2017, AJ, 154, 94 (APOGEE Overview Paper), York et al. 2000 (SDSS-I and SDSS-II Legacy technical description), Strauss et al. 2002 (SDSS Main Galaxy Sample Targetting), Frieman et al. 2008 (SDSS-II Supernova Survey technical description), Eisenstein, D.J., et al. 2001, AJ, 122, 2267 (LRG Sample Targetting), Richards, G.T., et al. 2002, AJ, 123, 2945 (Quasar Sample Targetting), (the Sloan Foundation 2.5-meter Telescope description) Gunn et al. 2006,

[2] Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey: Lintott, Chris J.; Schawinski, Kevin; Slosar, Anže; Land, Kate; Bamford, Steven; Thomas, Daniel; Raddick, M. Jordan; Nichol, Robert C.; Szalay, Alex; Andreescu, Dan; Murray, Phil; Vandenberg, Jan

[3] Galaxy Zoo 1: data release of morphological classifications for nearly 900 000 galaxies: Lintott, Chris; Schawinski, Kevin; Bamford, Steven; Slosar, Anå¾e; Land, Kate; Thomas, Daniel; Edmondson, Edd; Masters, Karen; Nichol, Robert C.; Raddick, M. Jordan; Szalay, Alex; Andreescu, Dan; Murray, Phil; Vandenberg, Jan

[4] [1808.04428] Deep learning of multi-element abundances from high-resolution spectroscopic data, Henry W. Leung, Jo Bovy. https://github.com/henrysky/astroNN

[5] B. Abolfathi *et al.*, The Fourteenth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the extended Baryon Oscillation Spectroscopic Survey and the second phase of the Apache Point Observatory Galactic Evolution Experiment, *Astrophysical Journal Supplement 235* (2018) 42.

[6] Multi-Spectral RGB-NIR Fig Classification Using Double-Channel CNN Jionghui Jiang1,2, Xi'an Feng1, Fen Liu1, Yingying Xu3, and Hui Huang 3, 1School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China 2Zhijiang College, Zhejiang University of Technology, Zhejiang 312030, China 3Computer Science Department, Wenzhou University, Wenzhou 325035, China

[7] Classification and Segmentation of Satellite Orthoimagery Using Convolutional Neural Networks by Martin Längkvist *OrcID, Andrey Kiselev OrcID, Marjan Alirezaie and Amy Loutfi Applied Autonomous Sensor Systems, Örebro University, Fakultetsgatan 1, Örebro 701 82, Sweden.